National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

AN EXAMINATION OF WISC-III FACTOR SCORES WITH A CLINICAL SAMPLE

BY

Laurie Fullerton

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Education.

IN

Counselling Psychology

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

Spring, 1994

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-11213-6

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Laurie Fullerton

TITLE OF THESIS: An Examination of the WISC-III

DEGREE: Master of Education

YEAR THIS DEGREE GRANTED: 1994

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Laurie Fullerton

Laurie Fullerton
100 Woodbine Road
Sherwood Park, Alberta
T8A 4A4

Date: March 21, 1994

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled An Examination of WISC-III Factor Scores With a Clinical Sample submitted by Laurie Fullerton in partial fulfillment of the requirements for the degree Masters of Education in Counselling.

Charles Norman

Dr. Charles Norman

Dr. Henry Janzen

Grace Malicky

Dr. Grace Malicky

November 3, 1993

DEDICATION

This thesis is dedicated to my family but especially to Remi, without whom I could not have finished graduate school or this thesis. He worked at my side, editing, advising, and giving me endless moral support. Thank-you for all the love and dedication you have given me in completing this project.

For my sons, Cameron and Derek John - thank you. They both did without "mom" on many nights and days and while they could not understand the magnitude and importance of this work they understood when I could not be with them.

For my mother, who financially and personally supported us throughout the good times and the bad - without her love and friendship I could not have completed graduate school or this manuscript.

For Dr. Charles Norman who had faith and conviction in my ability to do this work and gave of his time pouring over statistical data and endless rewrites with me.

For Tim, who always believed that I could complete this thesis and who worked with me in enhancing my self-esteem.

Finally, for Elizabeth, who loved and cared for all my family as if we were her children.

For all of you this thesis is dedicated. Thank you for giving me the courage and the support to carry on.

Abstract

The WISC-III was introduced in 1991. Ninety subjects, all referred for educational assessment, were given the WISC-III as part of a standard battery. Protocol results were entered into the Statistical Package for the Social Sciences, and a maximum-likelihood factor analysis was performed. Results indicate that the factor structure for a clinical sample results in a three factor loading, with no clear verbal-performance factors. Freedom from Distractibility and Processing Speed factors were not supported. Additionally, analyses of variance, using factor index scores, were performed on categories of exceptional children. Freedom from Distractibility and Processing Speed Index Scores did not differentiate between categories of exceptional children. Post-hoc analyses were also performed after the initial results were considered. It was hypothesized that the original sample, including a relatively large number of slow learners, may have skewed the results. Therefore, a second maximum-likelihood factor analysis was run, eliminating the slow learners from the sample. A second analysis of variance was also done, using only learning disabled and clinical non-categorical subjects. Results of this study indicate that factor scores of the WISC-III should not be used with exceptional children until further research is completed.

ACKNOWLEDGEMENTS

I would like to thank Dr. Henry Janzen who helped me in choosing this thesis topic and gave me the opportunity to utilize important data from the Education Clinic. Dr. Janzen also sat on my committee and provided valuable comments which helped me complete this project.

I gratefully acknowledge the staff at the Education Clinic for their efforts in helping me collect my data.

I would also like to thank Dr. Grace Malicky, who sat on my committee and provided suggestions for the improvement of my thesis.

Finally, I would like to thank the staff at the Sherwood Park Catholic School System for giving me the time and their support while I finished this project.

TABLE OF CONTENTS

Chapter	Title	Page
Chapter One	Introduction	1
Chapter Two	Review of the Literature	11
	Testing: Ancient	11
	19th Century: Pre-Binet	12
	1879-1903: Binet	15
	Post-Binet Reinterpretation	19
	Impact of the Army Testing Program	23
	Post-WWII: The Wechsler Scales	25
	WISC-R	29
	WISC-III	34
	Significant Changes of WISC-III	36
Chapter Three	Methodology	41
	Description of Subjects	41
	Procedure	41
Chapter Four	Results	46
	Post-Hoc Analyses	55
Chapter Five	Discussion	64
	Limitations of the Study	79
	Future Research Implications	79
	Personal Observations	80
	References Cited	83

LIST OF TABLES

Table	Title	Page
1	Placement of Subjects into Exceptional Groups	44
2	Full Scale IQ Scores & Distribution of Sample	46
3	Principal-Components Analysis	48
4	Rotated Factor Matrix for the Sample	50
5	Summary of Major Trends of Maximum-Likelihood Factor Analysis for the Total Clinical Sample	51
6	Distribution of Subjects based upon FSIQ and Achievement Scores	52
7	Analysis of Variance by Freedom from Distractibility Index-Category	54
8	Analysis of Variance by Processing Speed-Category	55
9	Maximum-Likelihood Factor Analysis-Sample without Slow Learners	58
10	Post-hoc Factor Matrix-Population without Slow Learners	59

...Tables Continued Next Page

Table	Title	Page
11	Summary of Subtest Loadings for Post-Hoc Maximum-Likelihood Factor Analysis	60
12	Post-hoc Analysis of Variance by Freedom from Distractibility Index with Learning Disabled and Normal Sample-Category	62
13	Post-hoc Analysis of Variance by Processing Speed Index with Learning Disabled and Clinical Non-Categorical Sample-Category	63

Chapter One

INTRODUCTION

Tests, of one kind or another, are used everyday in a variety of situations. Testing is used in business, in Universities, and in school districts. These tests can be informal, such as the ones teachers give in the classroom to determine the learning of a particular subject, or tests can be formal, standardized instruments which assess achievement, personality, or intelligence.

A review of the literature indicates that there is controversy regarding testing of children in general, and intelligence tests in particular. A large part of the criticism aimed at intelligence instruments is for the "use" of the scores (intelligence quotients referred to as "IQ") rather than the tests themselves. Many people believe that IQ test results are used for "labels" which may adversely impact the lives of many children. These labels, some believe, result in segregation or isolation from the "normal" population. Some people believe that IQ results are misinterpreted, leading to mistaken impressions of a person's "intelligence".

As a teacher and special educator, I have routinely used a variety of tests as part of an assessment. It is from the frequency in which I use intelligence tests that this thesis began: it is grounded research. I was well aware that noted authors such as Wechsler and

Sattler defend the intelligence test when it is correctly interpreted and used with a complete, individualized battery of instruments. However, I have had my doubts regarding the use of intelligence tests. At times I had seen parents and teachers misinterpret assessment results. I have also wondered about the usefulness of IQ scores in determining educational categories of exceptional children.

IQ scores are related to funding special education; however, government funding is a poor reason for putting children through the rigors of testing. In my opinion, the most important reason for giving tests is to provide information regarding a child's intellectual abilities so that educators may decide the most effective instructional strategies for each child.

Therefore, when the Psychological Corporation released the revised edition of the WISC-R (WISC-III) in 1991, I undertook the task of examining certain aspects of this test with a clinical sample. The WISC-R has been the most commonly used intelligence test to date. While IQ test results are commonly used in schools and industries, historically, intelligence testing is relatively new: the first IQ test was published only 88 years ago. A review of the literature reveals that it was not until Galton's work in the early 19th century that the study of individual differences was seriously considered. Up until the 18th century testing was given to determine eligibility for

the civil service. When measurement of individuals was undertaken, physical characteristics such as head size and arm span were used to assess intelligence.

With Galton's discovery of the mathematical/statistical procedures of correlation and regression to the mean, the field of psychometrics flourished. Scientists such as Cattell (1860-1944), Wundt (1879) and Pearson (1857-1936) continued to further develop the study of intelligence through the use of statistics; however, these scientists did not focus their studies on the development of an instrument with which to assess intelligence. It was Binet who is credited with creating the first intelligence test, which he devised in order to help "mentally defective" children.

Binet had a belief that all children had the potential to learn and that every student had the right to the best educational alternative. Binet contributed to the belief that special education could improve the learning of special needs children. Today special education models are built upon the fundamental philosophy which Binet advocated.

After 15 years of work, Binet introduced the first intelligence test in 1905. He produced age-graded tasks for children from the age of three to thirteen. Children were compared to their peers in order to differentiate between "normal" and "backward". The practice of

norm referenced tests continues today with exceptional children being compared to same age peers. Binet's research in test construction and theory are the cornerstone of modern intelligence tests.

The twentieth century saw American psychology influenced by behaviorism while the Europeans focused upon structural approaches reflected by the work of Jean Piaget. In the United States behaviorism was extended through psychometrics as it was considered the best methodology producing tangible educational results. Three American scientists are credited with refining the intelligence test (Binet's 1905 scale): Terman, Goddard and Yerkes. Terman's studies eventually resulted in the 1916 version of the "Stanford Revision and Extension of the Binet-Simon Scale" (Sattler, 1988). The Stanford-Binet is one of the two major IQ tests given today, the other being the Wechsler Scales. However, the introduction of intelligence tests created some mistrust in society.

The roots of fear and mistrust regarding IQ scores developed at the same time World War I began. Robert Yerkes translated the individual IQ test into a mass army testing device, the Army Alpha, including the advancement of the point-scale format. Commencing in 1917, over 1,000,000 men were tested with the Army Alpha and placement decisions, such as the difference between being an officer

or an enlisted man, were based upon test scores. It is important to emphasize that during this time the intelligence test relied heavily upon the skill of reading: this would have an important impact for immigrants and low socioeconomic status individuals between 1917 and 1939.

People from low economic backgrounds had limited access to schools resulting in poor reading and language skills. These people were discriminated against based upon the results of low IQ scores. As the emphasis was placed upon reading skills, many immigrants to America also scored low on intelligence tests. Immigration policies were set into place based upon test results; for example, people from Northern Europe were believed to be inferior and therefore, in 1912, the U.S. government placed quotas upon the number of these people allowed to immigrate.

Following World War I, reliability and validity became important statistical concepts and tests were refined during the period between World War 1 and World War II. In 1939, David Wechsler created the Performance Scale for use with the Army testing battery. This non-verbal scale would change the face of intelligence tests, as it de-emphasized the skill of reading. Wechsler believed that both the verbal and non-verbal tests were important; he therefore created a battery comprising of 11 subtests, many of

which still exist in revised formats today.

By 1975, Wechsler's scales had become the most popular IQ test used in the assessment of learning difficulties. This was due in part to Wechsler's de-emphasis upon reading and the strong validity and reliability of the test. Also in 1975, Kaufman reported the three factor solution for the WISC-R: Verbal Comprehension, Perceptual Organization and Freedom from Distractibility. Factor scores were increasingly used to generate hypotheses regarding special populations, eventually leading to the automatic use of factor scores when giving the WISC-R. In particular, low Freedom from Distractibility scores were linked with hyperactive children. Studies researching factor scores of exceptional children were numerous during the following decade (1975-1985).

In 1991, 25 years after the first Wechsler Scales were published, the Psychological Corporation introduced the WISC-III. Retaining 73% of the original items in the WISC-R (Sattler, 1992), this latest version has updated norms, test materials, and administration procedures. One of the goals of the WISC-III was to enhance the factor structure and consequently a new subtest was added, Symbol Search, which was to load on a new fourth factor, Processing Speed.

The WISC-III will be used as frequently as the WISC-R has been used in the past. Already it is the intelligence test of choice for

school boards in the province of Alberta. It will be used to assess the differences between exceptional and the "average" children. Given the history behind the IQ test, and the frequency with which it is used, I believed that an examination of the WISC-III and its factor structure would be relevant to my own work and that of others.

All statistical results of the WISC-III are primarily based upon the standardization sample: 2200 children, ages six to 16, from across the United States (WISC-III Manual, 1991). Jerome Sattler, who is considered a leader in the area of intelligence tests, has questioned the four factor structure of the WISC-III. He proposes that a three factor solution is more appropriate: Verbal Comprehension, Perceptual Organization, and Processing Speed. Sattler's exclusion of Freedom from Distractibility (FFD) is noteworthy given the amount of research linking FFD with hyperactive and learning disabled children. Steven Little (1992) also questions the use of the FFD index and the new factor, Processing Speed.

The amount of controversy generated regarding the use of the WISC-III factor scores led me to my first question whether, with a clinical population, a two, three or four factor loading would result. As factor scores are used for generating hypotheses regarding exceptional children, it followed that the factor structure for special populations should reflect that found for the standardization sample.

Furthermore, I wondered whether Processing Speed could differentiate amongst special population categories. Reviews of the WISC-III completed in 1992 questioned the validity of Processing Speed as a factor. Kaufman (1992) indicated that all children process material at different speeds, for a variety of reasons.

Thus the primary focus of this project was to examine the factor structure of the WISC-III using a clinical sample. Therefore, the sample was discrepantly different from the one used by Wechsler and Sattler. Ninety children, all referred for learning problems in school, were assessed at the University of Alberta Education Clinic. Unlike Wechsler's sample, demographic information was limited: variables such as social economic status and ethnic background were not available. Wherein the WISC-III standardization includes a small sample of special children, over 45% of the subjects in the present study were found to belong to a clearly defined exceptional category: gifted, slow learner, or learning disabled. The remainder of the sample consisted of children who were having difficulty in school, but these difficulties did not impact FSIQ and achievement scores. Other variables, such as behavioral, motivational, social or emotional difficulties may have affected their performance in school. This group, for the present study, were still considered "clinical", yet were identified as "non-categorical" as they

did not fit criteria (based upon test scores) for an exceptional group.

All subjects were given the WISC-III by trained graduate clinicians. Achievement tests given varied according to the referring problem. These results (WISC-III and achievement scores) were obtained from the clinic and placed into a database. Two procedures were initially performed: a maximum-likelihood factor analysis (varimax rotation) and analyses of variance. The factor procedures were identical to the those used by Wechsler and Sattler; the analyses of variance examined whether Processing Speed and Freedom from Distractibility could differentiate between categories of exceptional children.

Today, in some parts of Europe, the use of intelligence tests has been banned. Testing is devoted to criterion-based instruments, and each battery is individualized to the particular needs of the client. Given the doubts regarding IQ scores in general, along with the questionable validity of some aspects of the scales themselves, perhaps a thorough review of the use of intelligence tests, and assessments in general, is in order. As Lohman (1993) notes "...it is time to exorcise the notion that good assessments will always yield scores that permit individuals to be ranked, on one or several scales" (Lohman, 1993, p. 21).

When Buros, a noted reviewer, was asked to comment about

the advances in intelligence tests after 50 years, he replied: "Except for the tremendous advances in electronic scoring, analysis, and reporting of test results, we don't have a great deal to show..." (1977, p. 10). The following chapters chart the history of testing, and examine one aspect of the most widely used instrument in North America: the WISC-III. The results of this study indicate that further research needs to be done concerning at least one aspect of the WISC-III: the use of Factor Index Scores with clinical populations.

Chapter Two

REVIEW OF THE LITERATURE

Introduction

Modern psychological testing descended from civil service exams and university academic testing as well as "studies by European and American scientists on the measurement of individual differences in behavior" (DuBois, 1970, p. 3). It has taken centuries to develop the refined, standardized intelligence tests which are used today.

Testing: Ancient

As early as 2200 B.C., the Chinese emperor used tests to assess his civil servants: "After three examinations, officials were either promoted or dismissed" (Dubois, 1970, p. 3). By the year 1115 B.C. candidates were tested for abilities in the "six arts": music, archery, horsemanship, arithmetic, writing and the ceremonies and rites of public and private life (Eysenck & Kamin, 1981; Dubois, 1970). If an officer scored well on these examinations, he was given a government position.

Testing underwent many different changes in China throughout the centuries and in the year 1370 men who wanted to become public servants (mandarins) underwent two examinations: one in the

fall and one in the spring. Applicants were asked to quote and interpret the Confucian classics and testing lasted three days and nights (Watson,1988; Dubois, 1970). Testing was both written and oral in nature. Only 1 in 10 men were admitted to public office (Dubois, 1970, p. 5).

The Chinese methods of testing were admired by Europeans as they became involved in the Far East around the sixteenth century. Both the British and the French advocated the Chinese system of written examinations as a "means of entry into the public service" (Dubois, 1970, p. 6). The French used tests in 1791 to select civil servants and in 1833 examinations were given in London to select recruits for the Indian Civil Service. After the year 1800, written examinations were required for entrance to Oxford University. Thus the use of competitive examinations was a standard practice in England and France by the second decade of the 19th century (Love, 1990; Dubois, 1970).

19th Century Developments: Pre-Binet

In the middle of the 19th century there was a movement towards using tests for the study of individual differences (Watson, 1988; Dubois, 1970). The need to study individual abilities stemmed from the confusion and differences between "idiots and lunatics"(Shouksmith, 1970). Jean Esquirol (1772-1840) was the first

to make a clear distinction between the difference of being incapacitated and mentally ill (Shouksmith, 1970). In 1838 Esquirol noted that "idiots never developed their intellectual capacities, whereas mentally deranged persons lost abilities they once possessed" (Sattler, 1988, p. 38). His work focused upon physical characteristics including a study of speech. Esquirol described the "verbal characteristics at the various levels of idiocy, [and] although they were very crude, may have been the first effective mental tests" (Love, 1990, p. 23).

The work of Sir Francis Galton was influential in the testing movement. Galton provided two important statistical concepts which would stimulate the growth of psychological measurement in the twentieth century: regression to the mean and correlation (Sattler, 1988, p. 38). These statistics would, in the future, allow for the study of the relationship between intelligence test scores of parents and children, as well as the statistical relationship between any two variables (Love, 1990, Sattler, 1988).

Galton was interested in individual differences. In 1883 he published "Inquiries into Human Faculty and its Development" in which he considered the difficulties of measuring mental characteristics. Galton established a public clinic (1884) where "for a small fee a person could have a series of measurements taken and

recorded"(Dubois, 1970, p. 13). These included measurements of height, arm span, weight and ability levels of sight, memory, steadiness of hand and discrimination of color (Dubois, 1970, p. 13). Galton's focus on sensory functions led him to believe that "those with the highest intelligence should also have the best sensory discriminatory abilities" (Sattler, 1988 p. 38). As a result of Galton's work in attempting to create an intelligence test he is considered the "true father of the testing movement" (Love, 1990, p. 24).

James Cattell (1860-1944) added on to and standardized Galton's testing procedures, and gave them the name "mental tests" (Sattler, 1988). Cattell had studied with Wilhelm Wundt at Leipzig where in 1879 the first psychological laboratory had been built. (Dubois, 1970; Love, 1990). Cattell assisted Galton in his laboratory in the study of individual differences. Cattell believed that "psychology should rest on a foundation of measurement" (Love, 1990, p. 24), and stressed the importance of experimentation. From 1890-1900 Cattell created over 50 measurements which primarily assessed sensory and motor abilities. His studies were significant:

> Cattell did make a valuable contribution to psychology by taking the assessment of mental ability out of the field of abstract philosophy and showing that mental ability could be studied experimentally and practically. (Sattler, 1988, p. 39)

Karl Pearson (1857-1936) also worked with Galton and was a professor of mathematics at University College in London (Love, 1990, p. 24). Between the years of 1900 and 1904 Pearson discovered

> product-moment formula for r [...] multiple correlation, biserial r, correction of correlations for changes in range, and the chi square test. (Dubois, 1970, pp 14-15)

Pearson's work has important significance as the statistics that he created are the basis for the norm-referenced tests used today. Correlations are used extensively in modern intelligence tests to determine validity and reliability (Kaufman, 1979; Love, 1990). Pearson is referred to as the "father of statistics".

The work of men such as Galton, Cattell, and Pearson would significantly influence the testing movement in the twentieth century. Galton's work, in particular, would lead the French psychologist Alfred Binet (1857-1911) to his examination of higher mental processes and individual differences: by the year 1905 Binet would create the first "intelligence test" .

1879-1903: Binet and the European Influence

The period from 1880 to 1905 saw various psychologists striving to advance the field of psychology through the study of individual differences (Kaufman, 1979; Dubois, 1970; Love, 1990). This period was considered a "laboratory" phase in mental

measurement, and processes such as "sensation, attention, perception, association and memory" were studied in hopes of understanding intelligence (Sattler, 1988, p. 40). Whereas earlier testing focused upon social order and instruction, the period from 1880-1911 saw Binet and others striving to solve educational problems associated with individual differences.

Binet, with the assistance of his student, Theodore Simon, worked at the turn of the century to develop an instrument which would "detect and diagnose the intellectual deficiencies of the 'mentally defective' (Luther & Quarter, 1990, p. 1). Binet's methods were designed to

> make the distinction between normal and abnormal children that will permit commissions who decide on the admission of children into special schools to make an exact distinction so that only really abnormal children will be sent to special schools. (Wolf, 1973, p. 141)

Binet stressed that his goal was to improve the intelligence of the 'mentally defective' so that they could learn and grow through special education (Binet & Simon, 1916). Thus, Binet's test was considered both a "static measure of past learning ...and a flexible diagnostic tool which he used to understand learning difficulties" (Luther & Quarter, 1990, p. 15).

Binet, in order to design such an instrument, needed to

define 'natural intelligence' :

> ...in intelligence there is a fundamental faculty....This faculty is judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence. (Binet, 1916, pp. 43-44)

He began his studies with the simple observations of his two daughters, Madeleine and Alice, and he "concluded that simple tasks like those used by Galton did not discriminate between children and adults" (Kaufman, 1990, p. 5).

Binet and Simon departed radically from the work of Galton. It took 15 years of study to develop a test that would look at the "differences" in children and fit Binet's complex definition of intelligence (Linden & Linden, 1968). The first landmark scale, published in 1905 and then subsequently in 1908 and again in 1911 were age graded tasks for children ages three to thirteen (see Sattler, 1988; Luther & Quarter, 1990; & Kaufman, 1990). In the 1905 version, Binet and Simon discovered that their scale was useful for much more than classifying a child at one of the three levels of retardation: moron, imbecile, idiot (Matarazzo, 1972). Therefore, Binet extended his studies to include children outside of the range of scores between morons and idiots. The revised edition (1908) assessed learning problems by comparing children with their peers.

Binet's tests were "normed"; that is, the child was expected to pass a task that was included in his/her age group. Test items required memory, vocabulary, reasoning, co-ordination and comprehension. Binet, in essence, defined what was "normal". Normal was relative to one's age group:

> [Binet]...turned his attention to the practical problem of distinguishing the "backward" child from the "normal" child in the Paris schools. To do this he had to know what was a normal child. Difficult as this promised to be, it was a good deal easier than the attempt to define intelligence. For Binet concluded, quite logically, that the standard of a normal child of any particular age was something or other which an arbitrary percentage of children of that age could do.
>
> Having fixed a percentage which he would henceforth regard as "normal," he devoted himself to collecting questions, stunts, and puzzles of various sorts, hard ones and easy ones. At the end he settled upon fifty-four tests, ...all which together would test intelligence as a whole. (Lippmann, 1976, p. 5)

Binet's work and theoretical considerations continued, primarily with school-age children, up until his death in 1911, at which time he foresaw "numerous uses for his method in child development, in education, in medicine, and in longitudinal studies predicting different occupational histories for children of different intellectual potential"(Matarazzo, 1972, p. 42).

Binet's work had many important implications for the testing

movement. He emphasized and focused on the performance of a child relative to peers and "he was pioneering an approach to the diagnosis of learning difficulties and a philosophy of hope about the utility of education in improving learning" (Luthar & Quarter, 1990, p. 6). Binet's advances in test construction and theory can be considered the fundamental cornerstone of modern intelligence tests (Sattler, 1988, DuBois, 1970, Wolf, 1973). While Binet did not operationally define "intelligence", his set of tasks "survive to the present day in most tests of intelligence for children and adults" (Kaufman, 1990, p. 5). However, Binet's belief that the intelligence test looked at the "potential" of the child would be modified between 1920 and 1970 :

> They [Binet's successors in the intelligence test] also adopted Binet's type of test and use of age-based comparisons for measuring intelligence; but as we shall see they refuted Binet's philosophy and altered his purpose in important ways. (Luthar & Quarter, 1990, p. 6)

Post-Binet Reinterpretation

In the first half of the twentieth century American psychology was "dominated by behaviorism, the psychometric approach to intelligence[...]" (Watson, 1988, p. 29). Other European work, such as the structural approach of Piaget, was largely ignored, as it lacked "a rigorous experimental methodology" and did not yield tangible "educational benefits" (Watson, 1988, p. 29). The major translators of

Binet's scale in America were Henry Goddard at Vineland, Lewis Terman at Stanford, and Robert Yerkes at Harvard (Kamin, 1976).

In 1908 Goddard introduced Binet's 1905 version of the scale to the United States (Love, 1990; Sattler, 1988, Dubois, 1970). After minor revisions and standardizing the scale on 2,000 children, Goddard introduced the 1908 Binet-Simon scale (Goddard, 1910). The Binet-Simon scale was used strictly for the mentally retarded, and remained the most commonly used scale until Terman's work in 1916 (Linden & Linden, 1968).

Lewis Terman (1911) was interested in assessing gifted school children, and in 1912 revised the Binet-Simon scale using a number of additional subtests (Lippmann, 1976, p. 5). The 1916 standardized version became a major contribution in the evolution of psychometric testing and was known as the "Stanford Revision and Extension of the Binet-Simon Scale" (Love, 1990). In 1929, Terman started a longitudinal study of 1500 gifted children whose IQ scores were 140 or greater (Terman, 1925-1959). Terman's study of gifted children, and their descendants, lasted more than 60 years; the creation of the Stanford Revision was, in part, designed to facilitate this study. Terman's scale (1916) was based upon comprehensive, systematic research (Kaufman, 1990). The standardization techniques included a half year for training examiners, and all records were

recorded by Terman himself (Dubois, 1970). Terman's development of precise standardization procedures significantly contributed to the refinement of the intelligence test (Luthar & Quarter, 1990). Terman is also credited for his work on the concept of mental quotient (found by dividing mental age by chronological age). He developed this formula based upon the work of Stern (1914): "mental quotient" was replaced with the term intelligence quotient (IQ) in the 1916 Binet-Simon scale. "IQ" has since become an important variable in classifying children for special education purposes. The Stanford-Binet (SB), as it is now known, has been revised four times [1937,1960, 1972, 1986] (Swanson & Watson, 1982).

With the 1916 edition of the Stanford-Binet, intelligence testing "became firmly established in American schools and psychological clinics" (Dubois, 1970, p. 53). When the U.S. joined the war in 1917 the intelligence test was considered a way of classifying army recruits into ranks (Anastasi, 1988; Kaufman, 1990; Lippman, 1976; Linden & Linden, 1968). The task of revising the Binet scales into an army testing device was given to Robert Yerkes, President of the American Psychological Association (Kaufman, 1990). In 1917 Yerkes, along with Arthur Otis (Terman's student) led the way in developing the Army Alpha and Army Beta tests, the first to be group administered (Dubois, 1970; Kaufman, 1990; Sattler, 1988).

Consisting of ten subtests, the Army Alpha was the largest use of the intelligence test to date.

The point-scale approach of the Army Alpha Tests was an important contribution. Yerkes, along with Bridges and Hardwick, "published the Point Scale in 1915" (Sattler, 1988, p. 43). The point-scale methodology is "one that advocates selection of tests of specified important functions rather than a set of tasks that fluctuates greatly with age level and developmental stage "(Kaufman, 1990, p. 8). Age scale tests incorporate the belief that definitive forms of behavior appear at particular stages of development. In contrast, the point scale format emphasizes the correctness, quality, and speed of a response. Raw scores for each subtest are converted to standard scores, leading to an eventual overall score or scores for the subject (Love, 1990; Luther & Quarter, 1990). Subtests are chosen to measure specific functions. In essence, point scale tests measure the same aspects of behavior at every age (Kaufman, 1990). This approach to the design of intelligence tests is maintained today.

The standardization of the Army Alpha Test was impressive: 1,726,966 men were tested with "one of the five forms" of the test (Dubois, 1970, p. 66). The test combined Yerke's point-scale approach, as well as the kind of skills that Binet had put forward, such as "Directions or Commands, Practical Judgment, Arithmetical

Problems, and Information" (Kaufman, 1990, p. 9). Attention was given to psychometric properties of the test: reliability and validity. "Hundreds of correlation coefficients were computed, including a number of multiple correlations, being used for the first time in connection with psychological test results" (Dubois, 1970, p. 66). Army Alpha Test results provided personnel decisions for the army, and

> When the war ended, there were examining units in 35 camps, staffed with 120 officers and 350 enlisted men. (One of the enlisted men was David Wechsler, whose intelligence scale, first published two decades later, was to challenge both the Stanford-Binet and the nonverbal performance measures). (Dubois, 1970, p. 66)

Impact of the Army Testing Program: Social Ramifications

The impact of the Army testing program was that psychologists were encouraged to use testing in other areas: educationally, clinically, and vocationally. "Psychologists now saw that their methods for measuring individual differences could be refined and extended into new areas" (Dubois, 1970, p. 67). With the end of World War 1, the focus of testing turned towards identifying learning problems, particularly classifying the mentally retarded:

> With the Great War over, many psychologists returned to a focus on IQ testing as a means of childhood assessment...psychometric tools to subdivide retarded individuals into whatever number of categories [were] currently in vogue. (Kaufman, 1990, p. 7)

The intelligence tests in the first half of the twentieth century relied heavily upon schooling and the ability to read (Kamin, 1976); therefore, the immigrant and poor members of society were often found to be "inferior" (Luthar & Quarter, 1990). Testing was used to classify, segregate and discriminate children and adults:

> they [intelligence tests] all reflect the euphemisms, the homilies, and the morals which were, indeed, the stock and trade of Poor Richard's Almanac, Noah Webster's blue-back speller, and McGuffey's readers. The child who grew up in a home and attended a school where these things were in common usage stood in distinct advantage to the newly arrived immigrant child. At a time when over half the children in schools were either immigrants or children of immigrants, this movement represented discrimination in a massive way. (Kamin, 1976, p. 350-351)

"IQ" was used to explain why children could not do schoolwork, why certain races could not work at certain jobs, and in 1912 restricted the number of immigrants allowed into the United States (Eysenck & Kamin, 1981; Block & Dworkin 1976). Test results became a self-fulfilling prophecy: "not only that the lower class learned to believe in the system, but worse, through internalizing that set of beliefs, made it work" (Eysenck & Kamin, 1981; Kamin, 1976, p. 355). The system of labeling and segregating individuals according to IQ scores continued until 1975 when the passing of Public Law 94-142

guaranteed the right to equal education for all handicapped children, as well as disallowing any single score to be used for classification purposes (Sattler, 1988). Legislation such as PL94-142, combined with the refinement of the intelligence tests has limited cultural bias and discrimination associated with standardized testing (Kaufman, 1990; Luthar & Quarter, 1990).

Post WWII: The Wechsler Scales

The refinement and precision of intelligence tests continued with World War II. The Army Performance Scale Examination was developed prior to 1939 and this test "would become the tools-of-trade for clinical psychologists, school psychologists, and neuropsychologists into the twenty-first century" (Kaufman, 1990, p. 7). Originally designed to test army recruits who could not take the Alpha or Beta scales, it was composed of the following subtests: Picture Completion, Picture Arrangement, Digit Symbol and Manikin and Feature Profile (Object Assembly). Block Design was developed by Kohs in 1923; however, the rest of the Performance Scale was David Wechsler's creation:

> Wechsler's influential Performance Scale was added to the Army battery, [t]o prove conclusively that a man was weakminded and not merely indifferent or malingering". (Kaufman, 1990, p. 7)

Wechsler worked throughout the 1930's developing his scale,

pairing verbal tests that discriminated among children of different ages with nonverbal tasks (Dubois, 1970; Linden & Linden, 1968; Kaufman, 1990, p. 7). Wechsler was interested in creating a clinical instrument, and did his initial work while employed at the Bellevue Psychiatric Hospital in 1932. Utilizing a point-scale format, Wechsler compiled 11 different subtests to create the Wechsler-Bellevue Intelligence Scale Form 1 [1939](Sattler, 1988). The original 11 subtests, taken from a multitude of sources were as follows:

> Information, Comprehension, Arithmetic, Digit Span, Similarities, Vocabulary, Picture Completion, Picture Arrangement, Block Design, Coding. (Patterson, 1953, p. 11)

This first form of the Wechsler-Bellevue was for adults and adolescents only; however, these subtests have remained relatively the same for 40 years.

Wechsler believed that both the verbal and non-verbal scales were of equal importance. This was a primary distinction between Wechsler and psychologists before him (e.g. Terman, Goddard & Yerkes). Both verbal and non-verbal scales contributed to the overall IQ score (Kaufman, 1979; Patterson, 1953).

Another definitive feature of Wechsler's work was that he conceived the IQ test as being a clinical instrument which had a multitude of uses in both health and educational settings (Truch,

1989; Kaufman, 1979). Wechsler's search for a clinical tool was guided by his belief in the nature of intelligence; that is, "the aggregate of global capacity of the individual to act purposefully, to think rationally and to deal effectively with the environment" (Wechsler, 1958, p. 7). Wechsler also added that

> ...intelligence is not always adaptive, nor does it inevitably involve abstract reasoning. Intelligence is multifaceted as well as multi-determined. What it calls for is not a particular ability but an overall competency or global capacity, which in one way or another enables the sentient individual to comprehend the world and to deal effectively with its challenges. (Wechsler, 1981, p. 8)

For Wechsler an overall IQ was an indicator of general mental ability. Wechsler used the factorial theory of Carl Spearman (1863-1936), who had noted that early pioneers of testing such as Galton and Binet had been measuring one mental faculty, such as attention or memory (DuBois, 1970). Spearman advocated pooling test results, as he noticed that tests of ability "correlated" with one another:

> That means that people who did well on one test tended to do well on others. When one performs a complete statistical procedure called "factor analysis" on all these different tasks, then a 'common factor' emerges. This is called g. It represents what is common to all of the tasks. (Truch, 1989, p. 12)

Wechsler believed in the presence of a number of specific abilities "with an underlying general ability, or "g" factor" (Patterson, 1953). With this in mind, he devised his original Wechsler-Bellevue 1, and retained this theoretical approach (with more complex additions) in subsequent revisions.

With the end of WWII, Americans focused upon the widespread learning disabilities within the school system (Watson, A., 1988). Wechsler's scales differed in approach from the Stanford-Binet thus the popularity of his scales within the testing community took time to establish:

> Although the Wechsler scales did not initially surpass the Stanford-Binet in popularity, serving an apprenticeship to the master in the 1940s and 1950s, the WISC and the subsequent revision of the Wechsler-Bellevue, Form I (WAIS; Wechsler, 1955) triumphed in the 1960s.
>
> With the increasing stress on the psychoeducational assessment of learning disabilities in the 1960's and on neuropsychological evaluation in the 1970s, the Verbal-Performance (V-P) IQ discrepancies and subtest profiles yielded by Wechsler's scales were waiting and ready to overtake the one-score Binet. (Kaufman, 1990, p. 8)

Wechsler refined his 1946 scale (Wechsler-Bellevue Form II) to create the very successful WISC [Wechsler, 1949] (Kaufman 1990; Sattler, 1988). This version covered the ages of 5 to 15 instead of the original ages of 10 to 59 (Kaufman, 1990).

By 1970 Wechsler's scales were the most frequently used as the WISC was school related and psychologists could predict a child's future success based upon the IQ score:

> The IQ does not reflect a global summation of the brain's capabilities and is certainly not an index of genetic potential, but it does predict school achievement effectively. (Kaufman, 1979, p. 9)

Correlations with achievement were significant. Wechsler scales are now frequently used as an instrument to assess special populations:

> The WISC, of course, continues to figure prominently in all the problem areas of educational psychology and child development, still attracting attention as a vehicle for differentiating various ethnic groups and diagnostic categories, and defining samples. (Ward, 1988, p. 50)

The Wechsler Individual Scale of Achievement-Revised

By 1975, the intelligence test had become a very important instrument both educationally and clinically (Love, 1990). The WISC-R was "[currently] the most popular and well-researched psychometric instrument for the assessment of intellectual functioning" (Mueller, Mancini & Short, 1984, p. 300; Kaufman, 1980; Sattler, 1988). Reviews of the WISC-R indicated it as being

statistically reliable:

> Such features as the excellent reliability of its three Scale scores; the "adequate" reliable specific variance of most of its subtests...the relative stability and remarkable similarity of its factor composition across age...sex,...intelligence levels...ethnic/racial groupings and clinical diagnostic groupings. (Mueller, Mancini, & Short, 1984, p. 300)

The WISC-R covers ages ranging from 6-0 to 16-11 years and contains 12 subtests (Sattler, 1988; Truch, 1989; Love, 1990). The separate verbal and nonverbal scales have significantly contributed to its use in making educational diagnosis (Sattler, 1988; Kaufman, 1980).

There are certain patterns and scores in the WISC-R profile that have been used to study clinical populations. Individual subtest scores (high versus low), subtest scatter, and factorial grouping schema have been studied in order to make diagnostic decisions. There is no clear consensus among researchers, however, on the use of these patterns for classification or diagnosis.

A review of the literature indicates that no firm foundation exists for utilizing WISC-R subtest profiles or scatter analysis for making diagnostic decisions (Kaufman, 1976; Bannatyne, 1968; Mueller, Matheson & Short, 1983; Kramer, Henning-Stout, Ullman, & Schellenberg, 1987). In most cases, the scatter between the referred

population and the standardized group are almost equal. As Sexton & Street (1985) note, the WISC-R was useful for discriminating differences in intellectual functioning, but had no application for the diagnosis of specific education deficiencies.

Profile analysis of WISC-R protocols cannot provide treatment interventions; however, this analysis is useful for determining strengths and weaknesses of exceptional children. As yet it is not clear that profile analysis distinguishes any child from the normative group and, if used, profile analysis must be interpreted cautiously (Wechsler, 1974; Sattler, 1982; Mueller, Mancini & Short, 1984).

There is a large amount of literature concerning the use of factor analysis in differentiating special populations. Kaufman (1975), using WISC-R standardization data, identified a "tripartite factor structure consistent across all 11 age levels" (Rossini & Kowalski, 1986, p. 815). Kaufman identified the following three factors in the WISC-R (Kaufman, 1979, p. 22):

Verbal Comprehension	Perceptual Organization	Freedom from Distractibility
Information	Picture Completion	Arithmetic
Similarities	Picture Arrangement	Digit Span
Vocabulary	Block Design	Coding
Comprehension	Object Assembly	
	Mazes	

Kaufman noted that these three factors were found with both sexes

and all racial groups and these findings have been confirmed in other studies (Sattler, 1988). Rossini *et. al.* (1986) reported cross-validation among "normal and atypical groups, including children and adolescents with psychiatric disorders[...]mentally retarded youth,[...]and children with various neuropsychological disorders" (p. 815). A large number of research studies have been completed which "indicate that the factors found in these groups are generally similar to those found in the standardization group" (Sattler, 1988, p. 127).

One of the important applications of Kaufman's factor structure is to provide support for the construct validity of the Verbal and Performance Scales:

> The robust Verbal Comprehension and Perceptual Organization factors for all age groups and for a variety of supplementary samples of exceptional minority-group children indicate that the Verbal and Performance IQs reflect a child's performance on real and meaningful dimensions of mental ability. Thus the discrepancy between these IQs may well suggest important differences in the child's learning style and ability to handle different types of stimuli. (Kaufman, 1979, p. 23)

> The factor analytic results give strong empirical support to interpretation of the Verbal and Performance IQs as separately functioning entities in the WISC-R. (Sattler, 1988, p. 127)

The third factor, Freedom from Distractibility, has been the topic of much debate. While Sattler defends the name as "most

appropriate" as "the name focuses on the ability to concentrate or remain attentive" (p. 127), the construct validity of the third factor has been questioned (Rossini & Kowalski, 1986; Wielkiewicz, 1990).

A prevalent theme in the literature is the use of the third factor, Freedom from Distractibility, in identifying learning disabled children (Cadwell & English, 1983; Lombard & Riedel, 1978, Juliano et. al., 1988). Kavale & Forness (1984) reviewed 94 studies concerning the WISC-R and learning disabled children and found that the largest "effect" sizes were for Digit Span, Arithmetic, Coding, and Information (all but one third factor subtests). A pattern such as this, Wielkiewicz notes, could suggest a disruption in the cognitive process underlying the Third Factor (Wielkiewicz, 1990, Kaufman, 1979). Additionally, there has been studies linking third factor scores in identifying mentally retarded individuals (Groff & Hubble, 1982; VanHagen & Kaufman, 1975), and with children displaying Attention Deficit Hyperactivity Disorder (Sattler, 1988; Cohen, Becker & Campbell, 1990). However, the prevalent conclusion in the literature is "further research will be needed to clarify what is measured by the third WISC-R[...]factor" (Sattler, 1988, p. 174). Whether the third factor measures an intellective or behavioral variable was one of the reasons that the WISC-R was revised.

The reader who is further interested in the use, procedures and

administration of the WISC-R is directed to more comprehensive reviews dedicated to the exploration of these issues (see Sattler, 1982, 1988; Kaufman, 1990; 1975).

WISC-III

Eighty-six years after the introduction of the first intelligence test (Binet, 1905), the Psychological Corporation published the WISC-III (Wechsler, 1991). "Its theoretical underpinnings and structure remain virtually intact from the Wechsler Intelligence Scale for Children - Revised" (Little, 1992, p. 148). The refinement of measuring individual differences continues, and the WISC-III will be used frequently in assessing special populations (Little, 1992; Kaufman, 1992).

Seventy-three percent of the original WISC-R items remain in the latest edition (WISC-III Manual). The test provides current normative data and updated test materials, content, and administration procedures. The primary reason for revising the test was to update the norms (Sattler, 1992). Updating the norms periodically is essential, "otherwise, average IQ scores will gradually drift upward and give a deceptive picture of a child's abilities relative to others in the same age group" (Wechsler, 1991, p. 3). The standardization sample is nearly identical to 1988 census data:

> The WISC-III was standardized on 2,200 children, 100 boys and 100 girls in each of 11 age groups[...]The sample

> was stratified on age, race/ethnicity, geographic region, and parent education (used as a measure of socioeconomic status). For race/ethnic membership, children were classified as White, Black, Hispanic, or Other....(Sattler, 1992, p. 1031)

> As with the WISC-R, the WISC-III has sound statistical properties. Reliability coefficients for VIQ, PIQ, and FSIQ are all in the .90s and above .85 for the factor index scores. Individual subtest reliabilities are lower but generally acceptable. Test-retest coefficients indicate good stability over time. (Little, 1992, p. 151)

The validity of the WISC-III is unclear; however, "much of the research on the validity of the WISC-R probably applies to the WISC-III" (Sattler, 1992, p. 1040). Readers are directed to the Wechsler Intelligence Scale - III manual for a detailed description of other statistical properties of the test, as well as a complete description of individual subtests.

The WISC-III contains 13 subtests, 6 in the Verbal Scale and 7 in the Performance Scale. There are five subtests in each scale that are considered standard: Information, Similarities, Arithmetic, Vocabulary, and Comprehension in the Verbal Scale; Picture Completion, Coding, Picture Arrangement, Block Design, and Object Assembly in the Performance Scale. Supplemental tests include Digit Span in the Verbal Scale, and Symbol Search and Mazes in the Performance scale (Sattler, 1992). There are three composite scores

(Full Scale, Verbal, and Performance IQ) and four "optional factor-based index scores" (WISC-III Manual, p. vi). The changes from the WISC-R to the WISC-III are as follows:

1. Addition of the Symbol Search Subtest
2. Updated Materials
3. Factor -Based Index Scores
4. Improved Testing Materials
5. Coding Subtest
6. Minimization of Bias
7. Improved Floor and Ceiling
8. Changes in Administration and Scoring

(Sattler, 1992; Little, 1992; Wechsler, 1991)

A complete description of the changes found in the WISC-III can be found in the manual.

Significant Changes: Research Questions in the WISC-III

One of the most significant changes in the WISC-III is the four factor-based index scores (Wechsler, 1991). One of the reasons for updating the WISC-R was to further study the factor structure and the Psychological Corporation promotes an additional factor, Speed of Processing. The inclusion of a new subtest, symbol search, is intended to load on the fourth factor (Processing Speed). The four factor model is significant as it departs from the three factor model in the WISC-R,

and is already contraversial considering the WISC-III's recent introduction.

The four factors, according to Wechsler (1991) are as follows:

Factor I

Verbal Comprehension: Information, Similarities, Vocabulary and Comprehension

Factor II

Perceptual Organization: Picture Completion, Picture Arrangement, Block Design and Object Assembly

Factor III

Freedom from Distractibility: Arithmetic & Digit Span

Factor IV

Processing Speed: Coding and Symbol Search

(adapted from Wechsler, 1991)

Sattler (1992) and Little (1992) disagree with the four factor index model after performing maximum-likelihood factor analysis:

> I believe that the four-factor model is inappropriate. ..the so-called Freedom from Distractibility factor did not emerge in 4 of the 11 age groups[...]When a subtest has a loading below.30 on a factor, it suggests that the subtest fails to make a contribution to that factor[...]A four-factor solution never emerged in the principal components analysis. These results suggest that the Freedom from Distractibility factor on the WISC-III should be disregarded until there is further evidence to support its use. (Sattler, 1992, p. 1046)

The two main factors account for approximately 45% of

> the variance, yet Freedom from Distractibility accounts for only 2-3% of the variance and Perceptual Speed for only 4-5% of the variance. The inequity among the factors is therefore evident. The factor loadings definitely do not support the existence of the Freedom from Distractibility factor. (Little, 1992, p. 152)

> The manual provides very little information with regard to interpretation of the WISC-III. This may lead many individuals to accept the factor structure at face value or to continue using a procedure such as the one developed for the WISC-R by Kaufman (1979)[...]Specifically, Processing Speed should be interpreted cautiously and Freedom from Distractibility should be ignored. (Little, 1992, pp 152-153)

Symbol Search was added to the battery so that "abilities measured by the third factor could be better distinguished" (Wechsler, 1991, p. 5). However, Little notes that "this is an interesting rationale considering that with its inclusion the authors describe a four factor structure with Symbol Search on the fourth factor" (1992, p. 149). Wechsler states that "all factors were confirmed with high ability, low ability and clinical cases" (Wechsler, 1991, p. 185). The disagreement between Wechsler and Sattler regarding the number of factors found in the WISC-III subtests is important and merits further research.

The concept of processing speed in relation to special populations is another important issue arising from the WISC-III.

"The term *Processing Speed* describes the hypothesized ability underlying the factor for both item content (perceptual processing) and mental process (speed)" (Sattler, 1992, p. 1044). The Psychological Corporation has included bonus points for speed in the arithmetic subtest, and coding, symbol search and block design all take into account speed of response. The inclusion of bonus points for speed is already generating criticism. Kaufman notes that "giving bonus points for speed to preschool children seems silly from every developmental and common-sense perspective" (Kaufman, 1992, p.158). He adds, "there is great variability in the ability levels of individuals who solve problems at different rates of speed" (Ibid, p. 157). Gifted children often do not work quickly, but methodically at a problem and thus may have their scores lowered by this new bonus point system. It will be interesting to determine how processing speed differs within special populations.

As the WISC-III is an integral part of the assessment process, the question concerning a three or four factor structure is important. For the time being, most clinicians only use factor scores to generate hypotheses regarding each individual case. However, whether three factors are used or four will make a difference in the conclusions that may be reached regarding special populations and may impact further research. There is also little reference within the WISC-III

manual regarding the influence of processing speed as a differentiating factor with clinical subsamples.

Due to the importance and widespread use of the WISC-III, in both educational and clinical settings, the following research questions are proposed:

1. Would a factor analysis, using a clinical population, result in a two, three or four factor loadings?

2. Among different categories of normal and exceptional children, would scores on Processing Speed and Freedom from Distractibility (Indices) significantly reflect the difference between categories?

Chapter Three

METHODOLOGY

Description of Subjects

Subjects were 90 children between the ages of 6 years 0 months to 16 years 11 months, all referred to the University of Alberta Education Clinic for psychological assessment. All children were referred for some type of problem which affected school performance: learning, behavioral or emotional. Parents paid a minimal fee to have their children assessed. There were 32 females and 58 males. Assessments were done under the supervision of a chartered psychologist. All clinicians were graduate students in the Department of Educational Psychology.

All participants (or guardians and parents) signed consent forms agreeing that data could be used for research and teaching purposes on the condition that anonymity be maintained. Treatment of participants was in accordance with the ethical standards of the APA (Principal 9, Research with Human Participants).

Procedure

All testing procedures were established by the Director of Testing. Subjects were assessed with individualized test batteries, and all batteries included the Wechsler Intelligence Test for

Children-III (WISC-III). Standardized instructions and testing conditions were maintained for each subject.

Client files were loaned to the author. Information taken from the files included the following: name, date of birth, sex, reason for referral, grade level, handedness, WISC-III scores, achievement and personality scores (when given). No reanalysis of protocol data was done. Test scores were entered onto a spreadsheet and kept at the author's home.

Within the sample of 90 subjects, 42 students were classified as being "average"; however, this categorization was made only upon the basis of Full Scale IQ score and achievement data. This percentage of the sample is still considered "clinical" as these children were experiencing problems of a serious enough nature to be brought in for assessment. While test scores were in the average range, these children suffered from a number of social, emotional, behavioral or motivation problems which effect the way they perform in school. Therefore, these 42 children from the sample are called clinical non-categorical subjects, and are included in both the factor analysis and anovas.

In order to answer Question One, concerning the factor structure of the WISC-III with a clinical sample, the scores were translated for use in the Statistical Package for the Social Sciences

(SPSS). A maximum-likelihood factor analysis (varimax rotation) was performed using standard scores.

In order to answer Question Two, whether or not Freedom from Distractibility and Processing Speed differentiate groups of exceptional children, subjects were classified into four groups based upon Full Scale IQ, math and/or reading achievement scores. Four different achievement tests were used: Wide Range Achievement Test-Revised, Canada Quiet, Woodcock Johnson Achievement Battery and the Peabody Individual Achievement Test-Revised. The choice of the test given was made by each student clinican. An anova was performed to determine differences among the means of each of the four groups. A summary groups and criteria for placement is found in Table 1.

Table 1

Placement of Subjects into Exceptional Groups (FSIQ & Achievement Scores)

Category	FSIQ	Achievement Criteria (Math &/or Reading)
Clinical Non-Categorical	85-115	Achievement scores within 1 SD of the mean
Gifted	>115	Achievement scores > 1SD of the mean
Slow Learner	<85	Achievement scores <1 SD of the mean
Learning Disabled	85-143	Achievement 1 SD below FSIQ of the mean

Note. FSIQ = Full Scale IQ

Gifted students with achievement scores 1SD below their FSIQ score were included in the learning disabled category. Ninety subjects were placed into categories; there were three clinical categorical groups (slow learner, gifted, and learning disabled); the fourth group contained subjects whose IQ and achievement scores did not clarify the referal concern. Other factors, such as behavior, motivation, or emotional difficulties may account for difficulties in school. These subjects (n=42) are labeled "clinical non-categorical".

Three subjects from the sample did not have achievement scores and as such were not included. In cases where students did not have both reading and math achievement scores (as in the case of some subjects referred for learning disabilities), only one achievement score was used to determine group placement.

Chapter Four

RESULTS

There were 90 subjects between the ages of 6 years 0 months to 16 years 11 months. There were 32 female and 58 male subjects. Distribution of Full Scale IQ Scores is shown in Table 2.

Table 2

Full Scale IQ Scores and Distribution of Sample

		Full Scale IQ Score		
Category	n	M	SD	SEM
Average	62	99.0	7.66	.97
Gifted	16	124.0	7.35	2.44
Slow Learner	12	77.8	6.88	1.99
Null Cases	3			
Total Cases	90	100.65	15.0	1.60

Note. n = number of subjects. M = mean average full scale IQ score for the group. SD = standard deviation. SEM = standard error of measurement.

The sample does not replicate Wechsler's sample (1991), as it does not consist of a random, stratified group of children ages six to 16 and the distribution was not representative of the population in general. Information regarding the social economic status and the

ethnic background of the participants was not available. One of the purposes of the study was to determine whether scores from this clinical sample would result in a three or four factor loading (Question One). Standard scores were utilized in each analysis.

Question One

In order to answer Question One, a maximum-likelihood factor analysis was performed and the results indicate that a three-factor model best characterizes the WISC-III with this sample data, using criterion eigenvalues of 1 or greater (Eigenvalues refer to the proportion of variance that a given factor can account for; it is a mathematical value obtained from the factor analysis). Factor 1 accounted for 39.5% of the variance; Factor 2 for 12.4%, and factor 3 for 8.5% of the variance, with a cumulated percentage of 60.3% of the variance for the first three factors (see Table 3). All other factors had an eigenvalue <1, making the three factor model mathematically the most appropriate.

Table 3

Principal-Components Analysis

Factor	Eigenvalue	%of Variance	Cumulative %age of Variance
1	5.13158	39.5	39.5
2	1.60747	12.4	51.8
3	1.09926	8.5	60.3
4	.96930	7.5	67.8
5	.87777	6.8	74.5
6	.66647	5.1	79.6
7	.53037	4.1	83.7
8	.44743	3.4	87.2
9	.41205	3.2	90.3
10	.36900	2.8	93.2
11	.35112	2.7	95.9
12	.29133	2.2	98.1
13	.24685	1.9	100.0

For the total sample eigenvalues were below 1 for the fourth factor. None of the four hypothesized factors, found by Wechsler, emerged.

For this sample, three double loadings emerged: Information loaded on the first and third factors (.549 on Factor 1; .623 on Factor 3); Similarities also loaded on Factor 1 and 3 (.702 and .369 respectively); and Object Assembly loading on the first and second factors (.619 and .494). [When a subtest has a loading below .35 on a factor, it suggests that the subtest fails to make a significant

contribution to that factor] (Sattler, 1991). The three factor solution provided the cleanest loadings. Subtest loadings are shown in Table 4.

Nine subtests loaded on the first factor from the analysis. Of the nine, four were verbal subtests (Information, Similarities, Vocabulary and Comprehension); four were performance subtests (Picture Completion, Picture Arrangement, Block Design, and Object Assembly); and one was a subtest which commonly loads on the Freedom from Distractibility factor (Arithmetic). The second factor from the analysis loaded five subtests: three were performance subtests (Mazes, Block Design, Object Assembly) and two were processing speed subtests (Coding and Symbol Search). The third factor from the analysis loaded two verbal subtests (Information and Similarities) and two Freedom from Distractibility subtests (Arithmetic and Digit Span). A summary of subtest loadings is illustrated in Table 5.

Table 4

Rotated Factor Matrix for the Sample

Subtest	Factor 1	Factor 2	Factor 3
Picture Completion	.750	.216	.056
Information	.550	-.036	.623
Coding	.099	.797	-.136
Similarities	.703	.029	.370
Picture Arrangement	.650	.266	.031
Arithmetic	.574	.088	.534
Block Design	.583	.391	.318
Vocabulary	.795	.011	.205
Object Assembly	.620	.494	.016
Comprehension	.701	-.052	.036
Symbol Search	.300	.766	.101
Digit Span	.066	.041	.836
Mazes	-.195	.454	.308

Note. All calculations used scaled scores.
Loadings ≥.35 indicate contributions to the factor.

Table 5

Summary of Major Trends of Maximum-Likelihood Factor Analysis for the Total Clinical Sample

Subtest	Factor 1	Factor 2	Factor 3
Picture Completion	.750		
Information	.550		.623
Coding		.797	
Similarities	.703		.370
Picture Arrangement	.649		
Arithmetic	.574		.534
Block Design	.583	.391	
Vocabulary	.795		
Object Assembly	.620	.494	
Comprehension	.701		
Symbol Search		.766	
Digit Span			.836
Mazes		.454	

Question Two

As part of the clinical report, Factor Index scores for Freedom from Distractibility and Processing Speed were calculated and reported. In order to answer Question Two, regarding whether Freedom from Distractibility and Processing Speed differentiated exceptional populations, subjects were grouped according to FSIQ scores and achievement data (see chapter three). Table 6 illustrates the distribution in each group.

Table 6

Distribution of Subjects Based Upon FSIQ and Achievement Score

		Full Scale IQ Score		
Category	n	M	SD	SEM
Non-categorical	41	98.8	7.5	1.16
Gifted	7	125.1	9.75	3.69
Slow Learner	12	77.8	6.88	1.99
Learning Disabled	27	107.0	14.1	2.71
Null Cases	3			
Total Cases	90	100.65	15.0	1.60

Note. n = number of subjects. M = mean average full scale IQ score for the group. SD = standard deviation. SEM = standard error of measurement.

The clinical non-categorical group consisted of subjects whose FSIQ scores and achievement test results fell within the average range. However, these subjects are still considered as "clinical" as the referal problem was of a serious enough nature to prompt parents to obtain an assessment. A variety of other problems, which did not affect test scores, may have accounted for the problems experienced in school: social, emotional, motivational or behavior difficulties. Therefore, this group of subjects (n=42) were included in the anovas, being considered a clinical non-categorical group. The mean (FSIQ

scores) of the learning disabled category is high as gifted children with low achievement scores were included in this group.

Analysis of variance was performed utilizing Freedom from Distractibility Index Scores (FDI) and Processing Speed Index Scores (PSI). For both analyses the F values for FDI, F=.636. p=.636 and for PSI, F=.765, p=.765 did not approach significance. Freedom from Distractibility and Processing Speed Factor Scores did not differentiate between the samples of exceptional children. F scores are illustrated in Tables 7 and 8.

Table 7

Analysis of Variance by Freedom From Distractibility Index and Category

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F*
Main Effects	16727.571	3	5575.857	.570	.636
Category	16727.571	3	5575.857	.570	.636
Explained	16727.571	3	5575.857	.570	.636
Residual	801657.267	82	9776.308		
Total	818384.837	85	9628.057		

Note. 86 cases processed. 4 cases were missing.
*$p < .05$

Table 8

Analysis of Variance by Processing Speed and Category

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F*
Main Effects	11313.622	3	3771.207	.384	.765
Category	11313.622	3	3771.207	.384	.765
Explained	11313.622	3	3771.207	.384	.765
Residual	814485.895	83	9813.083		
Total	825799.517	86	9602.320		

Note. 87 Cases were processed. 3 cases were missing.
*$p < .05$

Post-Hoc Analyses

Index scores are commonly used for generating hypotheses regarding individual diagnosis; therefore, when the original factor analysis failed to confirm Wechsler's results, the sample used in the present study was reviewed. A post-hoc analysis was performed as an addendum to Question One as it was hypothesized that if the slow learners were removed, a different factor structure might emerge that was more typical of Wechsler's and Sattler's. Therefore, an additional maximum-likelihood factor analysis (varimax rotation) was performed because of the relatively large number of slow

learners (n = 12) in the original sample.

Post-hoc results indicate that, after excluding the slow learner group, a five factor model best characterizes the WISC-III with this population (criterion eigenvalues >1). Factor 1 accounted for 19.9% of the variance; Factor 2 for 18.4%; Factor 3 for 11.9%; Factor 4 for 9.0% and Factor 5 for 7.9% of the variance. This resulted in a cumulated percentage of 67.1% of the variance for the five factors. The sample (without the outliers) was not representative of either Wechsler or Sattler's positions.

In this analysis there was support for verbal and performance factors. On the first factor from the analysis six subtests loaded: four were performance subtests (Picture Completion, Picture Arrangement, Block Design, and Object Assembly); one was a verbal subtest (Comprehension) and one was a processing speed subtest (Symbol Search). The verbal subtest (Comprehension) negatively loaded on factor one. On the second factor from the analysis, six subtests loaded: three were verbal (Information, Similarities, and Vocabulary); two were Freedom from Distractibility subtests (Arithmetic and Digit Span) and one subtest from Processing Speed negatively loaded on the factor (Picture Completion). Factor three from the analysis loaded three performance subtests (Picture Completion (- loading), Coding, and Mazes), one verbal subtest

(Information), one processing speed subtest (Symbol Search) and one freedom from distractibility subtest (Digit Span). Factor four loaded three subtests: two verbal (Vocabulary and Comprehension) and one performance subtest (mazes). The fifth factor from this analysis loaded one performance subtest (Picture Arrangement) and one verbal subtest (vocabulary). The five factor solution (with the slow learners eliminated from the sample) provided the cleanest loadings (A summary of these results is illustrated in Table 9, 10, and 11).

Table 9

Maximum-Likelihood Factor Analysis
Sample Without Slow Learners

Factor	Eigenvalue	Percentage Variance	Cumulative Percentage
1	2.58888	19.9	19.9
2	2.38839	18.4	38.3
3	1.54487	11.9	50.2
4	1.17579	9.0	59.2
5	1.02717	7.9	67.1
6	.84378	6.5	73.6
7	.76737	5.9	79.5
8	.73161	5.6	85.1
9	.57257	4.4	89.6
10	.48611	3.7	93.3
11	.41189	3.2	96.5
12	.23788	1.8	98.3
13	.22099	1.7	100.0

Note. PC Extracted 5 Factors

TABLE 10

Post-Hoc Factor Matrix

Population Without Slow Learners

	Factor				
Variable	1	2	3	4	5
Picture Completion	.642	.160	-.408	.153	-.124
Information	.145	.654	.396	-.135	.103
Coding	.266	-.709	.376	-.052	.213
Similarities	-.022	.762	.138	.007	-.336
Picture Arrangement	.570	-.034	-.196	.250	.452
Arithmetic	.336	.586	-.223	-.162	.062
Block Design	.729	.196	.113	-.016	-.334
Vocabulary	.004	.437	.068	.563	.507
Object Assembly	.743	-.141	-.332	.026	-.002
Comprehension	-.441	.098	.003	.688	-.157
Symbol Search	.552	-.337	.514	.151	-.131
Digit Span	.118	.357	.593	-.258	.344
Mazes	.221	-.096	.479	.402	-.326

Note. All calculations used scaled scores.
Loadings ≥.35 indicate contributions to the factor.

Table 11

Summary of Subtest Loadings for Post-hoc Maximum-Likelihood Factor Analysis

	Factor				
Variable	1	2	3	4	5
Picture Completion	.642		-.408		
Information		.654	.396		
Coding		-.709	.376		
Similarities		.762			
Picture Arrangement	.570				.452
Arithmetic		.586			
Block Design	.729				
Vocabulary		.437		.563	.507
Object Assembly	.743				
Comprehension	-.441			.688	
Symbol Search	.552		.514		
Digit Span		.357	.593		
Mazes			.479	.402	

Note. All calculations used scaled scores.
Loadings ≥.35 indicate contributions to the factor.

A second post-hoc was performed as an addendum to Question Two. A prevalent theme in the literature is the use of the third factor, Freedom from Distractibility, in identifying learning disabled children [defined as those children whose achievement scores are at least one standard deviation below FSIQ] (Cadwell & English, 1983;

Lombard & Riedel, 1978, Juliano et. al., 1988). Given the emphasis upon learning disabled children and WISC-R Freedom from Distractibility scores, a second post-hoc analysis was performed utilizing the FDI and PSI factor index scores (as in the original study) for clinical non-categorical [achievement and FSIQ equivalent] versus learning disabled subjects. This was done so that the original findings (indicating that Processing Speed and Freedom from Distractibility did not differentiate between exceptional populations) could be confirmed. It was hypothesized that the gifted and slow learner scores [high and low end of the distribution] might have effected the original results [showing that PSI and FDI scores did not discriminate between exceptional samples].

For both analyses the F values for FDI, F=.701, p=.701 and for PSI, F=.365, p=.365 did not approach significance. The findings were the same as in the original study, indicating that Processing Speed and Freedom From Distractibility scores do not differentiate between exceptional populations. Analyses of variance for FDI and PSI are illustrated in Tables 12 and 13.

Table 12

Post-Hoc Analysis of Variance by FDI With Learning Disabled and Clinical Non-Categorical Sample

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F*
Main Effects	17.866	1	17.866	.149	.701
CAT	17.866	1	17.866	.149	.701
Explained	17.866	1	17.866	.149	.701
Residual	7788.642	65	119.825		
Total	7806.507	66	118.280		

Note. 67 cases were processed
*$p < .05$

Table 13

Post Hoc Analysis of Variance by PSI With Learning Disabled and Clinical Non-Categorical Sample

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F*
Main Effects	191.655	1	191.655	.834	.365
Cat	191.655	1	191.655	.834	.365
Explained	191.655	1	191.655	.834	.365
Residual	14943.807	65	229.905		
Total	15135.463	66	229.325		

Note. 67 cases were processed
*$p < .05$

Chapter Five

Discussion

As noted in chapter two, The WISC-III was only recently introduced (1991) and is widely used in the assessment of exceptional populations. The WISC-R was revised, in part, to "enhance the factor structure" (Wechsler, 1991). A new subtest, Symbol Search, was added to the standard battery which hypothetically loads on the fourth factor, Processing Speed. Thus far, two researchers have questioned the four factor solution proposed by Wechsler: Jerome Sattler (1992) and Steven Little (1992). In both cases, the researchers have been utilizing the standardization sample of 2200 children.

The aim of the present study was to examine the factor structure of the WISC-III utilizing a clinical population. An additional goal was to see if Processing Speed or Freedom from Distractibility would differentiate between groups of exceptional children (categorical and non-categorical). The clinical non-categorical group of children (n=42) had problems which did not affect FSIQ and achievement scores; behavioral, emotional, motivational and social difficulties may have affected performance at school. The sample used is different from Wechsler's normal random sample as it was

not representative of the general population.

The results indicate that factor scores can not be mathematically supported when considering an exceptional population. Most significant in the maximum-likelihood factor analysis was the absence of definitive verbal and performance factors. In the WISC-III manual it states "numerous factor analyses of WISC-R[...]typically showed two prominent factors, Verbal Comprehension and Perceptual Organization" (Wechsler, 1991, p. iv) and "the first two factors [in the WISC-III], Verbal Comprehension and Perceptual Organization are identical to those found in the WISC-R" (Wechsler, 1991, p. 187).

Kaufman (1975), using WISC-R standardization data identified a "tripartite factor structure consistent across all 11 age levels" (Rossini & Kowalski, 1986, p. 815). Kaufman noted that these three factors were found with both sexes and all racial groups and these findings have been confirmed in other studies (Sattler, 1988; Reschly, 1978). Rossini *et. al.* (1986) report cross-validation among "normal and atypical groups, including children and adolescents with psychiatric disorders[...]mentally retarded youth,[...]and children with various neuropsychological disorders" (p. 815). A large number of research studies have been completed which "indicate that the factors found in these groups are generally similar to those found in

the standardization group" (Sattler, 1988, p. 127). Wechsler (1991)notes the following:

> Numerous factor analyses of scores of clinical and school-based samples have indicated factor structures that closely align with the Verbal-Performance dichotomy. The Verbal-Performance structure was identified across amples varying in clinical and special-education classification[...](Wechsler, 1991, p. 181)

One of the important applications of Kaufman's factor structure is to provide support for the construct validity of the Verbal and Performance Scales:

> The robust Verbal Comprehension and Perceptual Organization factors for all age groups and for a variety of supplementary samples of exceptional minority-group children indicate that the Verbal and Performance IQs reflect a child's performance on real and meaningful dimensions of mental ability. Thus the discrepancy between these IQs may well suggest important differences in the child's learning style and ability to handle different types of stimuli. (Kaufman, 1979, p. 23).

> The factor analytic results give strong empirical support to interpretation of the Verbal and Performance IQs as separately functioning entities in the WISC-R. (Sattler, 1988, p. 127)

It is not clear in the WISC-III manual if standard or raw scores were used for the factor analysis, although it is assumed that standard scores were used. In the present study the use of raw scores presented a two factor loading, which was very

disproportionate. The use of raw scores would be more appropriate for an analysis of "g". Standard scores, therefore, were used with this clinical sample.

Little notes that "the two main factors account for approximately 45% of the variance" (1992) in the WISC-III data; the present analysis indicates that the first factor accounts for 39.5% and factor 2 for 12.4% of the variance. Nine subtests loaded on factor 1, compared to the four noted in the WISC-III manual. Of the nine subtests which loaded four were verbal (Information, Similarities, Vocabulary, and Comprehension), four were performance (Picture Completion, Picture Arrangement, Block Design, and Object Assembly), and one was a Freedom from Distractibility subtest (Arithmetic). Factor one could be labelled as "Verbal"; however, this factor is confounded by the presence of four additional subtests not traditionally thought of as reflecting verbal ability.

The second factor from the analysis loaded five subtests: three were performance subtests (Mazes, Block Design, Object Assembly) and two were processing speed subtests (Coding and Symbol Search). Wechsler's two strong Verbal-performance factors were not confirmed in the present analysis. Therefore, results from the clinical sample do not clearly confirm Wechsler's Verbal-Performance dichotomy, or Kaufman's three factor structure, nor does this analysis

support the findings of Little (1992) or Sattler (1992) who both find the presence of Verbal-Performance factors.

Due to the discrepancy between results of this study and Wechsler's finding of a four-factor model with a verbal-performance dichotomy, a post-hoc analysis was performed where slow learner scores were eliminated from the sample. With those scores removed from the sample, a five factor solution emerged, with a larger cumulative percentage of the variance accounted for. A strong verbal-performance factor structure, similar to Wechsler's and Sattler's is supported. However, this factor analysis is difficult to interpret and generally is not representative of either Wechsler's or Sattler's position.

The first factor from this analysis strongly loaded four performance subtests (Picture Completion, Picture Arrangement, Object Assembly, and Block Design). The second factor loaded three verbal subtests (Information, Similarities, and Vocabulary). Factors 3 (Picture Completion, Information, Digit Span, Coding, Symbol Search, Mazes); 4 (Vocabulary, Comprehension, Mazes), and 5 (Picture Arrangement, Vocabulary) do not represent any clear pattern. There were numerous double loadings in this analysis (Picture Completion, Information, Coding, Picture Arrangement, Comprehension, Symbol Search and Mazes). Negative loadings were noted for Picture

Completion, Coding, and Comprehension. This would suggest that some of the children in this sample may not have understood the questions on some of the subtests. The post-hoc analysis would confirm that the factor structure is questionable when used with clinical populations. It may be that learning disabled children have a different understanding, from that of the average population, when asked to complete certain tasks. Another hypothesis for why this factor structure differs from Wechsler's may be that these children (clinical sample) have gaps or uneven learning in their knowledge background which interferes with their perception of the subtests resulting in an uneven response pattern in the test itself.

A four factor model did not emerge using the clinical population, which differs from the findings of Wechsler (1991) who studied three groups other than the standardization sample (clinical, high-ability, and low-ability). Wechsler found that "[f]or the clinical group study...the four-factor model fit the data best for this clinical sample" (1991, p. 196). The absence of four factors in the present study would support Sattler who concludes the following:

> I believe that the four-factor model is inappropriate[...]the so-called Freedom from Distractibility factor did not emerge in 4 of the 11 age groups[...]When a subtest has a loading below .30 on a factor, it suggests that the subtest fails to make a contribution to that factor[...]A four-factor solution never emerged in the principal components analysis. (Sattler,

1992, p. 1046)

The three factor structure is difficult to interpret (Information, Object Assembly and Information have double loadings). Besides the absence of verbal and performance factors, Freedom from Distractibility and Processing Speed were not supported.

The lack of evidence for a third factor is noteworthy as Freedom from Distractibility (commonly noted as the third factor), has been the topic of debate. While Sattler defends the name as "most appropriate" as "the name focuses on the ability to concentrate or remain attentive" (Sattler, 1988, p. 127), the construct validity of the third factor has been questioned (Rossini & Kowalski, 1986; Wielkiewicz, 1990) in clinical practice. Factor Index Scores are used for enhancing individual diagnosis and "this use has been facilitated greatly by convenient tables for the rapid computation of factor deviation quotients in the familiar Wechsler metric [M=100, SD=15]" (Gutkin, 1982). One can assume that Factor Index Scores of the WISC-III will be used as frequently as WISC-R factor scores were used.

Freedom from Distractibility is a prevalent theme in the literature in the identification of learning disabled children (Cadwell & English, 1983; Lombard & Riedel, 1978, Juliano et. al., 1988). Kavale & Forness (1984) reviewed 94 studies concerning the WISC-R and learning disabled children and found that the largest "effect"

sizes were for Digit Span, Arithmetic, Coding, and Information (all but one third factor subtests). A pattern such as this, Wielkiewicz notes, could suggest a disruption in the cognitive process underlying the Third Factor (Wielkiewicz, 1990, Kaufman, 1979a). Additionally, there have been studies linking third factor scores in identifying mentally retarded individuals (Groff & Hubble, 1982; VanHagen & Kaufman, 1975), and with children displaying Attention Deficit Hyperactivity Disorder (Sattler, 1988; Cohen, Becker & Campbell, 1990). However, the prevalent conclusion in the literature is that "further research will be needed to clarify what is measured by the third WISC-R[...]factor" (Sattler, 1988, p. 174). Whether the third factor measures an intellective or behavioral variable was one of the reasons that the WISC-R was revised. .

In the present sample of exceptional children, Arithmetic (.53) and Digit Span (.83) strongly loaded on the third factor; however, Information had a highly loading (.62) on the same factor. The WISC-III manual indicates that only Arithmetic and Digit Span load on factor 3. A post-hoc analysis (which eliminated slow learners from the sample) confirmed that Wechsler's third factor, Freedom from Distractibility, was not supported. The results from the present study support Sattler, who notes that "Freedom from Distractibility[...]should be disregarded until there is further evidence

to support its use" (Sattler, 1992, p. 1046). In a study conducted by Little (1992), Arithmetic loaded strongly on factor 3, yet also loaded .41 on the Verbal Comprehension factor (factor 1). Little adds that "[i]nterpretation of this factor [Freedom from Distractibility]for anyone would not be recommended" (1992, p. 152). The results of the present study would support the cautionary comments made by Little .

There was also no support for the existence of Processing Speed. No loadings greater than .35 were found to substantiate a four factor model. Eigenvalues for factor 4 were less than one accounting for only 7.5% of the variance. Sattler proposes a three factor solution; however, he found support for Processing Speed as the third factor (Coding and Symbol Search). In the present study, however, Symbol Search loaded on factor 2 along with Coding, Object Assembly and Mazes. There is inequity among the factors which supports Little's analysis that "Processing Speed should be interpreted cautiously and Freedom from Distractibility should be ignored" (Little, 1992, pp. 152-153).

In attempting to answer why the results from Question One were so discrepant with Wechsler's findings, it was hypothesized that outlying scores may have altered the results. Therefore, a post-hoc analysis was performed which eliminated slow learners from the

sample. Results from this analysis showed that Symbol Search does not contribute to a factor called "Processing Speed". Symbol Search loads with Picture Completion, Digit Span and Mazes on a third factor. These findings are contrary to the Psychological Corporation's statement that symbol search was added to the battery so that "abilities measured by the third factor [Freedom from Distractibility] could be better distinguished" (Wechsler, 1991, p. 5). Little (1992) and Sattler (1992) note this is an interesting statement considering that, in the original sample of subjects, "Symbol Search" loads upon the fourth factor, Processing Speed. With an exceptional population Symbol Search does not load on the third factor nor does it statistically contribute to a fourth factor, "Processing Speed". This would refute Wechsler who states that "all factors were confirmed with high ability, low ability and clinical" cases (Wechsler, 1991, p. 185).

The inclusion of "Processing Speed" in the WISC-III was noteworthy as one could hypothesize that the speed with which a child processes information would differentiate exceptional learners. Freedom from Distractibility has been utilized for years in generating hypotheses regarding learning disabled and attention-deficit children and it might follow that Processing Speed would also enhance classification of exceptional children.

Results from the present study showed that processing speed did not differentiate between four categories of exceptional children. An analysis of variance indicated that F scores for PS of .765 were not significant. It was hypothesized that the gifted and slow learner scores may have affected the results. Therefore, a post-hoc analysis of variance was performed using only learning disabled and clinical non-categorical (those whose referring problem was not clear from FSIQ and achievement scores) subject scores. F scores for PS of .365 were not significant: processing speed did not differentiate between the two groups.

Kaufman notes, in speaking about gifted students, that

> The main point is that there is great variability in the ability levels of individuals who solve problems at different rates of speed (Kaufman, 1979b), and that variability is undoubtedly due largely to non-intellective factors such as poor motor coordination or a reflective cognitive style. Whenever personality, behavioral, or motor variables enter in a child's scores on an intelligence test, then the role assumed by "pure" intelligence becomes less. The children who are most affected by the intrusion of "irrelevant" variables are invariably the ones who are referred for gifted assessment. (Kaufman, 1992, p. 157)

Children will perform tasks at various speed for various reasons. The personalities of each child may dictate how he or she approaches a task: diligently or haphazardly. Affective variables may also influence speed of response on any particular day: if a child is tired,

depressed, or anxious his speed of response may vary significantly from the norm. Therefore, to assume that speed of response may be used to differentiate children is erroneous.

Given the large amount of research in identifying learning disabled children using Freedom from Distractibility Index (FDI) scores, an additional analysis of variance was performed using FDI index scores. Index scores were compared with the three categories and one non-categorical group of clinically referred children (non-categorical, gifted, slow learner, and learning disabled) and results indicate that FDI scores did not discriminate among the four groups of children. A post-hoc analysis was conducted comparing only learning disabled and non-categorical childrens' scores. It should be noted that the present sample used a diverse clinical population. In categorizing the learning disabled subjects, gifted students who scored 1 standard deviation or lower on achievement tests were included in the "learning disabled" category. The inclusion of gifted learning disabled children increased the mean (of FSIQ scores) for this group. However, post-hoc analysis repeated the finding that Freedom from Distractibility did not discriminate between LD and clinical non-categorical subjects (whose FSIQ and achievement were average).

A review of the literature found that third factor scores were

often used with the WISC-R in identifying learning disabled children. Learning Disabled children often had low scores on the third factor, and "[are] usually interpreted when the three subtest scores are similar in magnitude and one of them deviates significantly from the mean of Verbal or Performance subtest scores..." (Grossman, 1985; Kaufman, 1979). However, as early as 1978 Gutkin had noted that the reliability of the Distractibility factor was too low for routine clinical interpretation (Gutkin, 1982). Wielkiewicz (1990) also notes that "many other hypotheses may account for low scores [on the third factor] (p. 91). The results of the present study would indicate that FDI scores should not be used to discriminate between exceptional populations.

The reliability of test scores is one of the fundamental principles of psychometrics. Before an instrument can be used to make diagnostic and classification decisions, it must be proved valid and reliable. The WISC-III will be used as extensively as the WISC-R has been used in the past. The major subtests and IQ scores found in the WISC-III are statistically reliable and valid. Wechsler has stated that Factor Index Scores are optional, and "should never [be] used as independent estimates of a child's Full Scale IQ" (Sattler, 1992, p. 1048). However, Index Scores are often used as a means of generating hypotheses when making decisions regarding exceptional

students. Sattler states that "they [factor based index scores] should simply be used in generating hypotheses about the child's performance" (1988, p. 174). The results of the present study would not support the use of factor-based index scores for gathering information regarding an exceptional student:

> The manual provides very little information with regard to interpretation of the WISC-III. This may lead many individuals to accept the factor structure at face value or to continue using a procedure such as the one developed for the WISC-R by Kaufman (1979). One should be extremely cautious in taking either of these approaches. As was previously mentioned, the factor structure does not appear to be as neat and clean as the promotional material sent out by the publisher may indicate....Finally, one should avoid using any interpretation system for which data are not available. (Little, 1992, p. 153)

The present study has illuminated an area of concern. Freedom from Distractibility factor scores have been used for the past 18 years, since Kaufman discovered the three factor structure in 1975. It was believed that the three factors "[gave] strong empirical support to interpretation of the Verbal and Performance IQs as separately functioning entities in the WISC-R" (Sattler, 1988, p. 27). The three factors were statistically confirmed with both sexes and all racial groups, using the original standardization data (Rossini et. al, 1986; Sattler, 1988; Reschly, 1978). Kaufman noted that

> The robust Verbal Comprehension and Perceptual Organization factors...for a variety of supplementary

> samples of exceptional minority-group children indicate that the Verbal and Performance IQs reflect a child's performance on real and meaningful dimension of mental ability." (Kaufman, 1979, p. 23)

Results from the present study do not confirm the existence of verbal-performance factors. Considering that the sample consisted primarily of exceptional children, these results indicate that using a technique such as the split between verbal and performance factor scores should be the focus of continued research. Even hypotheses generated about exceptional children utilizing such a technique (the V-P split) might be misleading, if not wrong.

The third factor, so-called Freedom from Distractibility, has also been used increasingly to identify learning disabled and hyperactive children. The debate surrounding the use of this factor has gained momentum in recent years (see Wielkiewicz, 1990; Sattler, 1988; Kaufman, 1990). As noted earlier in this chapter, Sattler has questioned the existence of Freedom from Distractibility, as has Little (1992). The results of this study would suggest that this factor does not exist when examining a group of clinically referred subjects. Processing Speed also has no statistical relevance in differentiating categories of exceptional children.

Limitations of the Study

There are two primary limitations with the present study.

First, the number of examiners could not be controlled. This affects the administration, scoring, and interpretation of test results (see Slate & Jones, 1990). Secondly, the sample size could have been larger. With a greater number of exceptional children, it would increase the confidence in the statistical findings.

Future Research Implications

Due to the findings that the Factor Index Scores are not significantly supported, it would be useful to examine a larger sample, which would include age range categories. In particular, the nature of the childrens' performance, using a larger clinical population should be examined on specific subtests that load on factor three and four (Freedom from Distractibility and Processing Speed). This would help to explain whether or not childrens' perceptions of these two subtests is, in some way, affecting test scores.

Personal Observations

A review of the history of the intelligence test shows that IQ scores traditionally lead to labels and special placements. However, when Binet conceived the intelligence test more than 88 years ago, his fundamental principle was that IQ tests could improve the quality of a child's education, and that the score would provide

better treatment for a diagnosis of learning difficulties. Today, with the emphasis upon funding, an IQ score leads to placement in special education: test results do not provide treatment of specific learning problems. As Lohman (1993) notes, "scores from existing tests of fluid abilities [intelligence tests] have few instructional implications" (p. 21).

Within the past 20 years the focus of testing has centered upon learning disabilities:

> Although many people have proposed that the WISC-R could be used for treatments related to learning disabilities and neuropsychological functioning, for example, there is ample evidence arguing against this. (Little, 1992, p. 153)

Considering that IQ tests are usually given when there is a problem with learning (exceptional children), and that thousands of trained psychologists use these tests everyday, the validity of all components of the instrument is of paramount importance.

An IQ score can have both positive and negative consequences for many children. The faith we place in these scores must be based upon statistically sound properties; otherwise, standardized testing may see the rebirth of an area with no substantiation. Additionally, we may be placing exceptional children at further risk. Automatic usage of the factor scores may take us back to shades of the old

auditory and visual processing tests used in the 1960's and 70's which incorrectly identified learning disabled children. The WISC-III has a great many properties that are outstanding and the recent revisions had eliminated some of the criticisms noted with the WISC-R. However, as the data indicate, the performance of these exceptional children on some levels replicates the findings of Wechsler, but overall the results are not clear. It would seem that exceptional children may be lost in the normative sample used by Wechsler and Sattler; additionally, disabled children may not understand some of the subtests. Until further research disproves the results of this study, factor scores should not be used with special populations.

Today, in Alberta, the practice remains not to use factor scores when writing an assessment. In conversations with colleagues I was told that they are, however, used as confirmation of the three IQ scores or are used when considering large discrepancies between verbal and performance test results. The results of this study lend credibility to the practice of omitting factor scores when considering IQ results. As our assessment report leads to placements and labels, using factor scores with exceptional children with no empirical support would be unethical. Also, utilizing these scores in the assessment of exceptional children could place them further at risk

than they already are.

References Cited

Anastasi, Anne (1988). Psychological Testing (6th ed). New York: Macmillan.

Bannatyne, A. (1968). Diagnosing learning disabilities and writing remedial prescriptions. Journal of Learning Disabilities, 4, 272-274.

Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. L'Annee Psychologique, 11, 191-244.[1]

Binet, A., & Simon, T. (1916). The development of intelligence in children (E.S. Kit, Trans.). Baltimore: Williams & Wilkins.

Block, N.J. & Dworkin, G. (Eds) (1976). The IQ Controversey. New York: Random House, Inc.

Buros, O.K. (1977). Mental Measurements Yearbook. University of Nebraska: Buros Institute of Mental Measurements.

Cadwell, J., & English, F. (1983, April). The effect of learning disabilities on the factor structure of the WISC-R: A test of factorial invariance. Paper presented at the meetings of the American Educational Research Association, Montreal.

Cohen, M., Becker, M., & Campbell, R. (1990). Relationships Among Four Methods of Assessment of Children With Attention Deficit-Hyperactivity Disorder. Journal of School Psychology, 28, 189-202.

DuBois, Philip H. (1970). A History of Psychological Testing. Boston: Allyn and Bacon Inc.

Eysenck, H. J. & Kamin, Leon (1981). The Intelligence Controversey.

[1] Binet's french reference is commonly cited in the literature (e.g., Sattler, 1988); however, no English translation is available. I am unable to read French and used this reference only to denote authorship.

United States: John Wiley & Sons, Inc.

Goddard, H. H. (1910). A measuring scale of intelligence. Training School, 6, 146-155.

Groff, M., & Hubble, L. (1982). WISC-R factor structures of younger and older youth with low IQs. Journal of Consulting and Clinical Psychology, 50, 148-149.

Grossman, F. M. (1985). Interpreting clinically derived WISC-R subtest groupings: A statistical approach. Journal of Psychoeducational Assessment, 3, 89-96.

Gutkin, T.B. (1982). WISC-R deviation quotients vs. tradition IQs: An examination of the standardization sample and some implications for score interpretation. Journal of Clinical Psychology, 38, 179-182.

Juliano, J.M., Haddad, F.A., & Carroll, J.L. (1988). Three year stability of WISC-R factor scores for black and white, female and male children classified as learning-disabled. Journal of School Psychology, 26, 317-325.

Kamin, Leon J. (1976). Heredity, Intelligence, Politics, and Psychology: II. In N.J. Block & Gerald Dworkin (Eds.) The IQ Controversey (pp. 374-382). New York: Pantheon Books.

Kaufman, A. (1975). Factor Analyses of the WISC-R at eleven age levels between 6.5 and 16.5 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kaufman, A. (1980). Issues in psychological assessment: Interpreting the WISC-R intelligently. In B. Lagey & A. Kazdin (eds.,) Advances in child psychology. New York: Plenum.

Kaufman, Alan S. (1976). A New Approach to the Interpretation of Test Scatter on the WISC-R. Journal of Learning Disabilities, 9 (3), 33-168.

Kaufman, Alan S. (1979). WISC-R research: Implications for interpretation. School Psychology Digest, 8, 5-27.

Kaufman, Alan S. (1979a). Intelligence Testing with the WISC-R. New York: Wiley-Interscience.

Kaufman, Alan S. (1990). Assessing Adolescent and Adult Intelligence. Boston: Allyn and Bacon, Inc.

Kaufman, Alan S. (1992). Evaluation of the WISC-III and WPPSI-R for Gifted Children. Roeper-Review, 14 (3), 154-158.

Kavale, K.A., & Forness, S.R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? Learning Disability Quarterly, 7, 136-156.

Kramer, J.J., Henning-Stout, M., Ullman, D.P., & Schellenberg, R.P. (1987). The Viability of Scatter Analysis on the WISC-R and the SBIS: Examining A Vestige. Journal of Psychoeducational Assessment, 5, 37-47.

Linden, K.W. & Linden, J.D. (1968). Modern mental measurement: A historical perspective. Boston: Houghton Mifflin.

Lippman, W. (1976). The Lippman-Terman Debate: Mental Age of Americans. In N.J. Block & Gerald Dworkin (Eds.) The IQ Controversey (pp. 4-13). New York: Pantheon Books.

Little, Steven (1992). The WISC-III: Everything Old is New Again. School Psychology Quarterly, 7 (2), 148-154.

Lohman, David F. (1993). Teaching and Testing to Develop Fluid Abilities. Educational Researcher, 22 (7), 12-23.

Lombard, T.J., & Riedel, R.G. (1978). An analysis of the factor structure of the WISC-R and the effect of color on the Coding subtest. Psychology in the Schools, 15, 176-179.

Love, Harold. D. (1990). Assessment of Intelligence and Development of Infants and Young Children. Illinois: Charles C. Thomas.

Luther Michael & Quarter, Jack (1990). The Genie in the Lamp: Intelligence Testing Reconsidered. Ontario: Captus Press.

Matarazzo, J.D. (1972). Wechsler's measurement and appraisal of adult intelligence. (5th ed.). Baltimore: Williams & Wilkins.

Mueller, H. H., Mancini, G. J., & Short, R. H. (1984). An Evaluation of the Diagnostic Efficiency of the Wechsler Intelligence Scale for Children-Revised. The Alberta Journal of Educational Research, XXX (4), 299-310.

Mueller, H., Matheson, D., & Short, R. (1983). Bannatyne-recategorized WISC-R patterns of mentally retarded, learning disabled, normal and intellectually superior children: A meta-analysis. The Mental Retardation and Learning Disability Bulletin, 11 (2), 60-78.

Patterson, C. H. (1953). The Wechsler-Bellevue Scales: A Guide for Counselors. Springfield, Illinois: Thomas.

Reschley, D. (1978). WISC-R factor structures among Anglos, Blacks, Chicanos, and native American Papagos. Journal of Consulting & Clinical Psychology, 46, 417.

Rossini, E.D., & Kowalski, J.M. (1986). WISC-R Factor Reliability at 11 Age Levels. Journal of Clinical Psychology, 42 (5), 815-817.

Sattler, J.M. (1974). Assessment of Children's Intelligence (rev). Philadelphia: Saunders.

Sattler, J. M. (1982). Assessment of children's intelligence and special abilities (2nd ed). Boston: Allyn & Bacon.

Sattler, Jerome (1988). Assessment of Children (3rd ed). San Diego: Jerome M. Sattler.

Sattler, Jerome (1992). Assessment of Children WISC-III AND WPSSI-R Supplement. San Diego: Jerome M. Sattler.

Sexton, Larry & Street, Sue (1985). A Comparison of WISC-R Score Profiles of Students Placed in Special Education and those Referred, but not Placed. Psychology in the Schools, 22, 142-145.

Shouksmith, G. (1970). Intelligence, creativity and cognitive style. New York: Wiley.

Slate, J. R., & Jones, C. H. (1990). Student Error in Administering the WISC-R: Identifying Problem Areas. Measurement and Evaluation in Counseling and Development, 23, 137-140.

Swanson, H. Lee & Watson, Billy L. (1982). Educational and Psychological Assessment of Exception Children. St. Louis: The C.V. Mosby Company.

Terman, L. (1916). The measurement of intelligence. Boston: Houghton Mifflin.

Terman, L.M. (1925-1959). Genetic Studies of Genius (Vols 1-5). Standford CA: Stanford University Press.

Truch, Steve (1989). The WISC-R Companion. Calgary: Foothills Educational Materials.

Van Hagen, J., & Kaufman, A.S. (1975). Factor analysis of the WISC-R for a group of mentally retarded children and adolescents. Journal of Consulting and Clinical Psychology, 43, 661-667.

Ward, James (1988). Developments in Intelligence Testing and its Application: A Personal Viewpoint. In Alan Watson (Ed.), Intelligence: Controversy & Change (pp. 44-60). Australia: The Australian Council for Educational Research Limited.

Watson, Alan J. (Ed) (1988). Intelligence: Controversy & Change. Australia: Australian Council for Educational Research.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised. San Antonio: The Psychological Corporation.

Wechsler, D. (1955). Wechsler Adult Intelligence Scale. NY: Psychological Corporation.

Wechsler, D. (1974). Selected Papers of David Wechsler. New York: Academic Press.

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). San Antonio, TX: The Psychological Corporation.

Wechsler, D.(1958) The measurement and appraisal of adult intelligence (4th ed). Baltimore: Williams & Wilkins.

Wechsler, D. (1949). Manual for the Wechsler Intelligence Scale for Children. New York: Psychological Corporation.

Wechsler, D., (1974). Manual for the Wechsler Intelligence Scale for Children-Revised (WISC-R). San Antonio, TX: The Psychological Corporation.

Wielkiewicz, R.M. (1990). Interpreting low scores on the WISC-R third factor: It's more than distractibility. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 91-97.

Wolf, Theta H. (1973). Alfred Binet. Chicago: The University of Chicago Press.